# Deep Learning From Crowdsourced Labels: Coupled Cross-Entropy Minimization, Identifiability, and Regularization

Shahana Ibrahim, Tri Nguyen, Xiao Fu

ICLR 2023, Virtual Talk


Oregon State
University

# Labeled Data

One of the pillars of AI revolution is ***massive amount of labeled data***

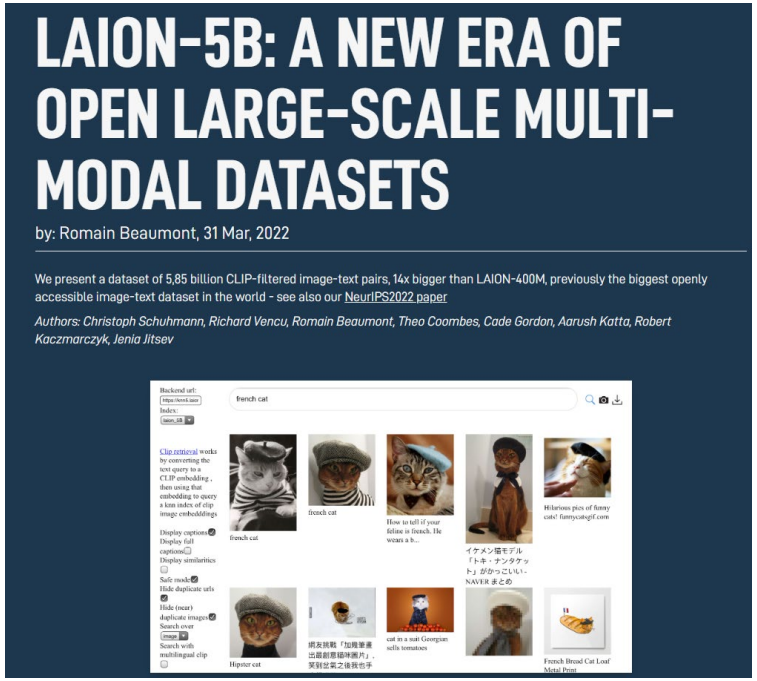

Amazon team taps millions of Alexa interactions to reduce NLP error rate

KYLE WIGGERS @KYLE_L_WIGGERS

JANUARY 22, 2019 6:59 AM

Amazon VP of devices David Limp at a September 2018 event at Amazon headquarters in Seattle, Washington.

Image Credit: Khari Johnson / VentureBeat



LAION-5B: A NEW ERA OF OPEN LARGE-SCALE MULTI-MODAL DATASETS

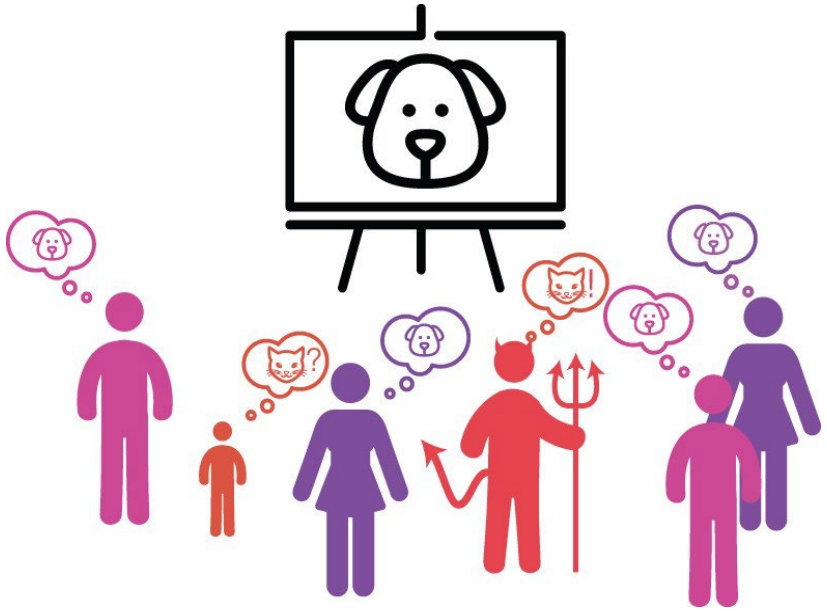
by: Romain Beaumont, 31 Mar, 2022

We present a dataset of 5,85 billion CLIP-filtered image-text pairs, 14x bigger than LAION-400M, previously the biggest openly accessible image-text dataset in the world - see also our NeurIPS2022 paper

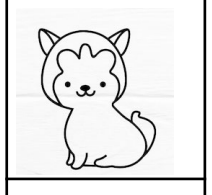
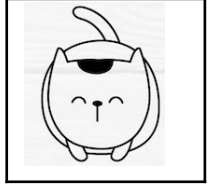
*Authors: Christoph Schuhmann, Richard Vencu, Romain Beaumont, Theo Coombes, Cade Gordon, Aarush Katta, Robert Kaczmarczyk, Jenia Jitsev*

https://laion.ai/blog/laion-5b/

# Crowdsourcing – Using Wisdom of the Crowd

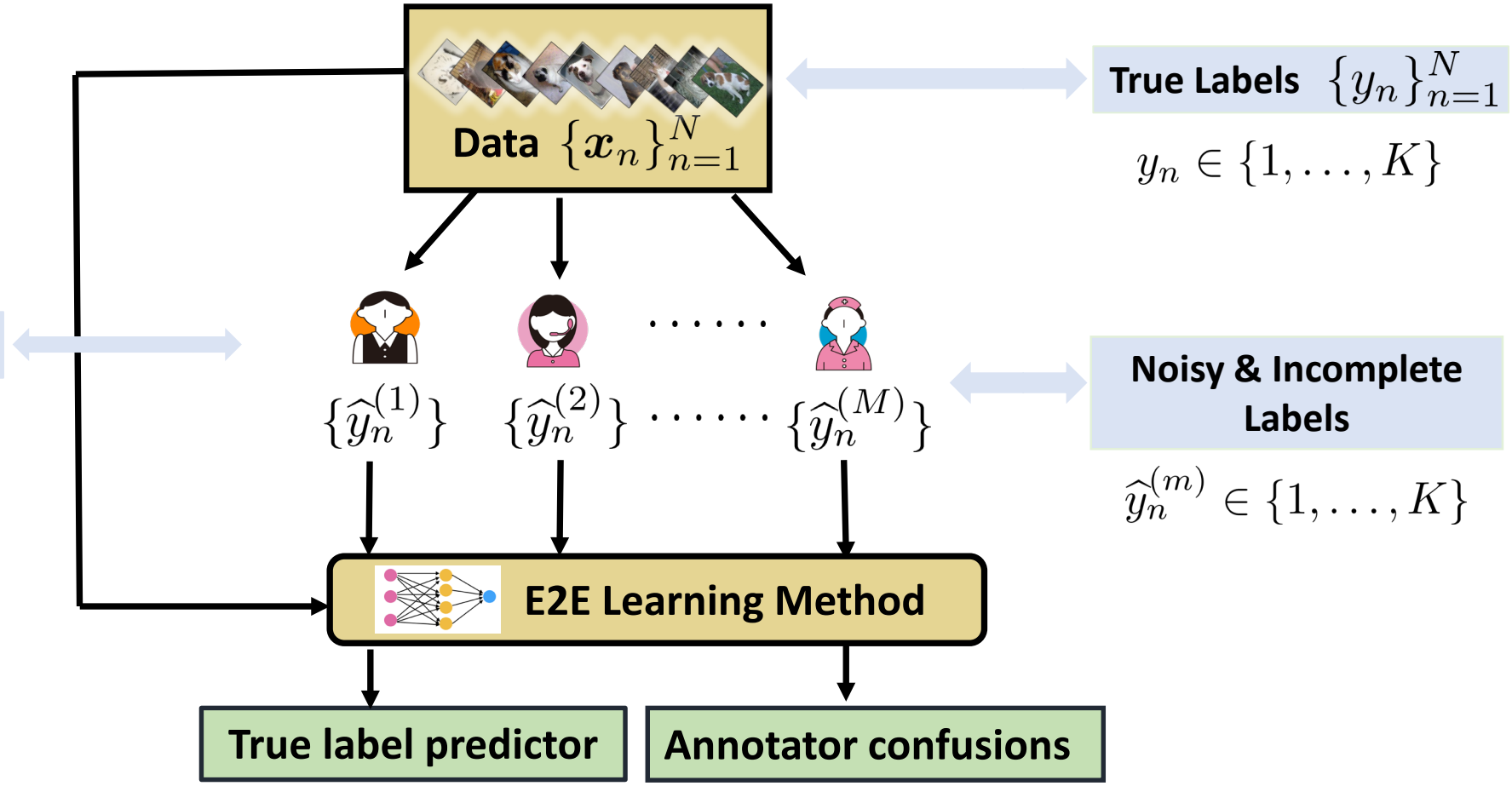


***Each data item is labeled by multiple, often non-expert, annotators***

| Data | Labels | | | |
|---|---|---|---|---|
|  | Cat | ? | ..... | Dog |
|  | Dog | Cat | ..... | ? |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ |
|  | ? | ? | ..... | Dog |

***Noisy & Incomplete***

# End-To-End Learning for Crowdsourced Labels




[Raykar et al., 2010]
[Khetan et al., 2018]
[Rodrigues & Pereira, 2018]
[Tanno et al., 2019]
[Li et al., 2020]
[Chen et al., 2020]
[Chu et al., 2021]
[Wei et al., 2022]

# Noisy Label Generation Model

$$\Pr(\widehat{y}_n^{(m)} = k|\boldsymbol{x}_n) = \sum_{k'=1}^{K} \Pr(\widehat{y}_n^{(m)} = k|y_n = k')\Pr(y_n = k'|\boldsymbol{x}_n)$$

**Prob. Of $m$th annotator's response given $\boldsymbol{x}_n$**

$\boldsymbol{p}_n^{(m)}$

**Confusion Matrix**

$\boldsymbol{A}_m$

**True label predictor**

$\boldsymbol{f}(\boldsymbol{x}_n)$

$$\boldsymbol{p}_n^{(m)} = \boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n), \forall m, n$$

$$\widehat{y}_n^{(m)} \sim \text{categorical}(\boldsymbol{p}_n^{(m)})$$

**Goal : learn $\boldsymbol{f}$ and $\boldsymbol{A}_m, \forall m$**

**Assumption: Annotator confusions are not data dependent**

# Coupled Cross Entropy Minimization (CCEM)

**The most popular E2E learning criterion [Rodrigues & Pereira, 2018]**

$$\underset{\boldsymbol{f},\{\boldsymbol{A}_m\}}{\text{minimize}} \quad \frac{1}{|\mathcal{S}|} \sum_{(m,n)\in\mathcal{S}} \text{CE}(\boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n), \widehat{y}_n^{(m)})$$

$$\text{subject to} \quad \boldsymbol{f} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

[Rodrigues & Pereira, 2018]
[Tanno et al., 2019]
[Li et al., 2020]
[Chen et al., 2020]
[Chu et al., 2021]
[Wei et al., 2022]

**Indices of observed labels**

**Neural network function class**

**Constrained set**
$\{\boldsymbol{A} \in \mathbb{R}^{K\times K} | \boldsymbol{A} \geq \boldsymbol{0}, \boldsymbol{1}^\top \boldsymbol{A} = \boldsymbol{1}^\top\}$

**Can CCEM learn the true label predictor and the true annotator confusions?**



F. Rodrigues and F. Pereira. ``Deep learning from crowds". Proceedings of the AAAI Conference on Artificial Intelligence, 2018.

## *Analysis Result*: CCEM correctly learns the true confusions and the true classifier,

**under the assumptions**

1. ***Anchor point condition*** → **For each class, there is a data sample belonging to that class with prob. close to 1**
2. ***Class expert condition*** → **For each class, there is an expert which can predict that class correctly with prob. close to 1**

**Often, its hard to hold the two conditions together**

# Proposed Learning Criteria

✓ ***Anchor point condition***

✗ ***Class expert condition***

→ **If we have more data, but no experts to label**

***GeoCrowdNet (F)***

$$\underset{\boldsymbol{f},\{\boldsymbol{A}_m\}}{\text{minimize}} \quad \frac{1}{|\mathcal{S}|} \sum_{(m,n)\in\mathcal{S}} \mathsf{CE}(\boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n), \widehat{y}_n^{(m)}) - \lambda \log\det \boldsymbol{F}\boldsymbol{F}^\top$$

$$\text{subject to} \quad \boldsymbol{f} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

**Regularization term**

$$\begin{bmatrix} \boldsymbol{p}_1^{(1)} & \cdots & \boldsymbol{p}_N^{(1)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{p}_1^{(M)} & \cdots & \boldsymbol{p}_N^{(M)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{f}(\boldsymbol{x}_1) & \cdots & \boldsymbol{f}(\boldsymbol{x}_N) \end{bmatrix}$$

$\boldsymbol{P}$ $\quad$ $\boldsymbol{W}$ $\quad$ $\boldsymbol{F}$

**Underlying NMF model**

# Proposed Learning Criteria

✗ ***Anchor point condition***

✓ ***Class expert condition***

→ **If we have less data, but there are experts to label**

***GeoCrowdNet (W)***

$$\underset{\boldsymbol{f},\{\boldsymbol{A}_m\}}{\text{minimize}} \quad \frac{1}{|\mathcal{S}|} \sum_{(m,n)\in\mathcal{S}} \mathsf{CE}(\boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n), \widehat{y}_n^{(m)}) - \lambda \log\det \boldsymbol{W}^\top \boldsymbol{W}$$

$$\text{subject to} \quad \boldsymbol{f} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

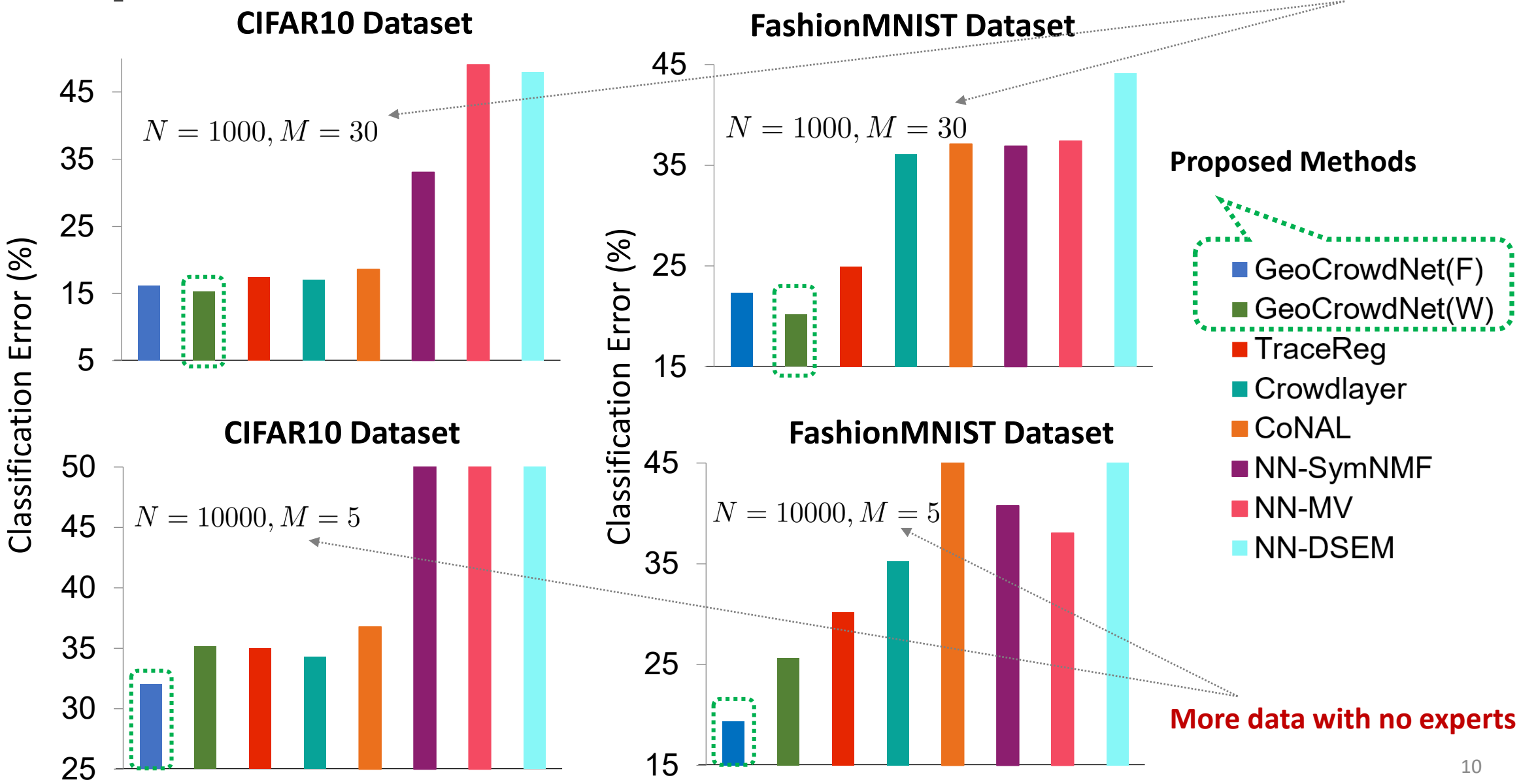
**Regularization term**

$$\begin{bmatrix} \boldsymbol{p}_1^{(1)} & \cdots & \boldsymbol{p}_N^{(1)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{p}_1^{(M)} & \cdots & \boldsymbol{p}_N^{(M)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{f}(\boldsymbol{x}_1) & \cdots & \boldsymbol{f}(\boldsymbol{x}_N) \end{bmatrix}$$

$\boldsymbol{P}$ $\quad$ $\boldsymbol{W}$ $\quad$ $\boldsymbol{F}$

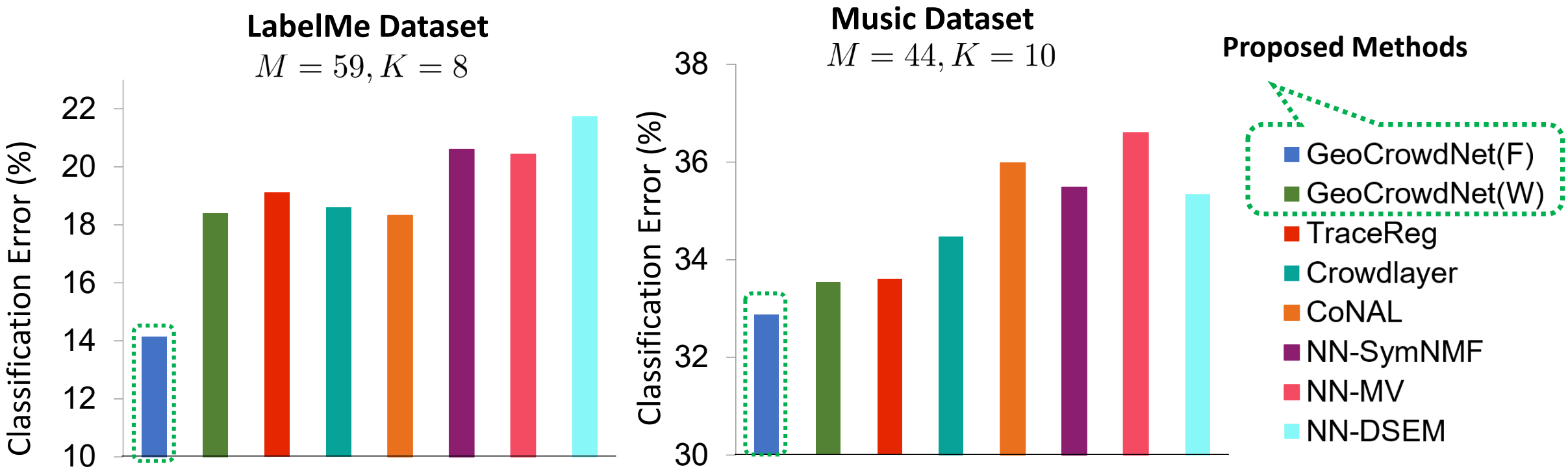
**Underlying NMF model**

# Empirical Results



# Empirical Results

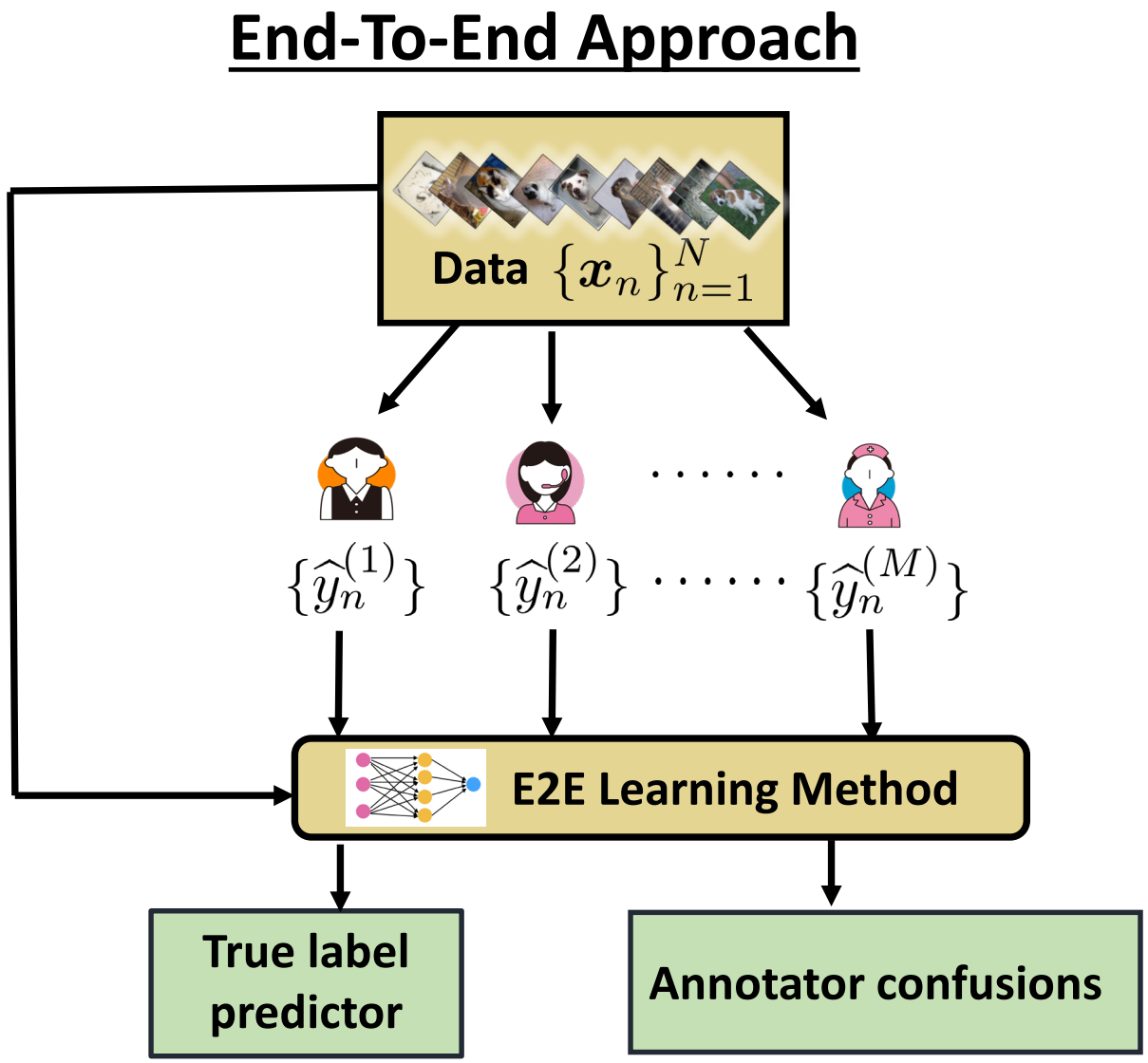
**Noisy labels from amazon mechanical turk workers**




**The criterion designed for ``no experts case" shows edge in practice**

# Key Takeaways

- Established deeper understanding to the challenging E2E learning problem
- Designed learning criteria with enhanced performance under practical scenarios, **e.g., no expert annotators**



# Thank You